Iran. J. Insur. Res., 13(3): Pages: 275-286, Summer 2024, Serial #49

Iranian Journal of Insurance Research
(IJIR)

Homepage: https://ijir.irc.ac.ir/?lang=en

Iranian Insurance Research Center (IIRC)

ORIGINAL RESEARCH PAPER

Estimation of micro-level claim reserving using copula function

S. Shams[1,*], M. Esna-Ashari [2], M. Piadeh Kouhsar[1]

[1] *Department of Statistics, Faculty of Mathematical Sciences, Alzahra University, Tehran, Iran*

[2] *Department of Property and Casualty Insurance Research, Iranian Insurance Research Center, Tehran, Iran*

ARTICLE INFO

Article History:
Received 24 July 2023
Revised 08 November 2023
Accepted 18 December 2023

Keywords:
Censored data
Copula function
Dependence structure
Reserved claims

*Corresponding Author:
Email: *s.shams@alzahra.ac.ir*
Phone: +9821 89562100
ORCID: 0000-0002-8831-3372

DOI: 10.22056/ijir.2024.03.06

ABSTRACT

BACKGROUND AND OBJECTIVES: The estimate of reserved claims is based on the prediction of the final amount of claims that have not yet been settled and which the insurer has undertaken to pay. The micro-level approach estimates the amount of each unsettled claim separately. In this context, the time taken to settle a claim is an important variable, as large claims usually take longer to settle and may not be paid in the relevant financial period. Non-payment of claims may be due to a lack of timely reporting, a high volume of cases or legal complications. Therefore, there may also be censored data. In this article, we use the copula function approach to model the dependency structure of the settlement duration and claim amount variables.
METHODS: In this study, the amount of each unpaid claim is estimated separately using a micro-level approach. For this purpose, the duration of the settlement of each claim is considered as a variable that depends on the amount of the claim. Based on the modeling of the dependency structure between the claim amount and the settlement duration using the copula function and using the general characteristics of the claims as predictor variables, each unpaid claim is estimated. The marginal distributions of the claim amount and the settlement period based on covariates explain the stochastic behavior of each of them and are modeled separately. In the copula function approach, the dependence structure of the variables can be considered separately from their marginal distributions. By choosing or constructing an acceptable and appropriate copula function with the dependence structure of the claim amount and the duration of its settlement, and taking into account the particular characteristics and conditions of inflation, the amount of an unpaid claim can be estimated more accurately. In this study, the accuracy of the estimation and the validity of the proposed model are evaluated using simulations.
FINDINGS: The proposed method is implemented for the data collection of one of the insurance companies regarding the cases of the employer's professional liability claims in 8 years, from March 2013 to March 2021, which includes unpaid claims. In order to evaluate the error in estimating the claim amount with this method in this data set, 1000 samples are taken from the data whose claim amount is known and settled and each time a number similar to the actual number is censored, the known values are censored and their amount is estimated. Then the estimation error is calculated using the criteria mentioned in the article. And finally, it can be seen that the results have a good accuracy rate.
CONCLUSION: In the article, it can be seen that the selected copula function has acceptable results for the estimation of reserved claims.

نشریه علمی پژوهشنامه بیمه

سایت نشریه: https://ijir.irc.ac.ir/?lang=fa

مقاله علمی

برآورد ذخیرهٔ خسارت‌های معوق در سطح خرد با استفاده از تابع مفصل

صدیقه شمس[1,*]، مریم اثنی‌عشری[2]، محبوبه پیاده کوهسار[1]

[1] گروه آمار، دانشکده علوم ریاضی، دانشگاه الزهرا(س)، تهران، ایران

[2] گروه پژوهشی بیمه‌های اموال و مسئولیت، پژوهشکده بیمه، تهران، ایران

اطلاعات مقاله

تاریخ های مقاله:

تاریخ دریافت: ۰۲ مرداد ۱۴۰۲

تاریخ داوری: ۱۷ آبان ۱۴۰۲

تاریخ پذیرش: ۲۷ آذر ۱۴۰۲

کلمات کلیدی:

تابع مفصل

داده‌های سانسور شده

ذخیره خسارت‌های معوق

ساختار وابستگی

*نویسنده مسئول:

ایمیل: s.shams@alzahra.ac.ir

تلفن: ۰۰ ۸۹۵۶۲۱ ۹۸۲۱+

ORCID: 0000-0002-8831-3372

DOI: 10.22056/ijir.2024.03.06

چکیده:

پیشینه و اهداف: برآورد ذخیرهٔ خسارت‌های معوق براساس پیش‌بینی مبلغ نهایی خسارت‌هایی است که هنوز تسویه نشده است و بیمه‌گر متعهد به پرداخت آن است. در رویکرد سطح خرد، میزان هر خسارت تسویه‌نشده به‌طور جداگانه برآورد می‌شود. در این راستا، مدت‌زمان تسویهٔ هر خسارت متغیر مهمی است، زیرا تسویهٔ خسارات بزرگ معمولا بیشتر طول می‌کشد و ممکن است در دورهٔ مالی مربوطه پرداخت نشود. تسویه نشدن مطالبات ممکن است به‌دلیل گزارش نشدن به‌موقع، حجم بالای پرونده‌ها و یا عوارض قانونی باشد. ازاین‌رو، ممکن است داده‌های سانسورشده نیز وجود داشته باشند. ما در این مقاله از رویکرد تابع مفصل برای مدل‌سازی ساختار وابستگی متغیرهای مدت‌زمان پرداخت خسارت و مبلغ خسارت، استفاده می‌کنیم.

روش‌شناسی: در این تحقیق، با رویکرد سطح خرد، میزان هر خسارت تسویه‌نشده به‌طور جداگانه برآورد می‌شود. برای این منظور، مدت‌زمان تسویهٔ هر خسارت به‌عنوان متغیر وابسته به میزان خسارت در نظر گرفته می‌شود. براساس مدل‌سازی ساختار وابستگی بین مبلغ خسارت و مدت تسویه با استفاده از تابع مفصل، و همچنین با استفاده از ویژگی‌های کلی خسارت‌ها به‌عنوان متغیرهای پیش‌بینی‌کننده، هر خسارت تسویه‌نشده برآورد می‌شود. مدل‌سازی جداگانهٔ توزیع‌های حاشیه‌ای مبلغ خسارت و مدت تسویه براساس متغیرهای کمکی، رفتار تصادفی هریک از آنها را توضیح می‌دهد. در رویکرد تابع مفصل ساختار وابستگی متغیرها را می‌توان جدا از توزیع‌های حاشیه‌ای آنها در نظر گرفت. با انتخاب یا ساخت تابع مفصل قابل قبول و مناسب با ساختار وابستگی مبلغ خسارت و مدت‌زمان تسویهٔ آن، و لحاظ کردن ویژگی‌های خاص و شرایط تورم، می‌توان میزان هرخسارت پرداخت‌نشده‌ای را با دقت بیشتری تخمین زد. در این تحقیق از شبیه‌سازی برای ارزیابی صحت برآورد و اعتبار مدل پیشنهادی استفاده می‌شود.

یافته‌ها: روش پیشنهادی برای مجموعه داده‌های یکی از شرکت‌های بیمه مربوط به موارد اعلامی ناشی از مسئولیت حرفه‌ای کارفرما در ۸ سال که شامل مطالبات پرداخت‌نشده نیز هستند، اجرا می‌شود. برای ارزیابی خطای برآورد میزان ذخیرهٔ معوق با این روش در این مجموعه داده، از بین داده‌هایی که مبلغ خسارت آنها معلوم است و تسویه شده‌اند به تعداد ۱۰۰۰ بار نمونه‌گیری شده و هر بار تعدادی مشابه تعداد واقعی سانسورشده، مقادیر معلوم سانسور می‌شوند و ذخیرهٔ معوق آنها برآورد می‌شود. سپس خطای برآورد با معیارهایی که در این مقاله آمده، محاسبه شده است. در نهایت مشاهده می‌شود که یافته‌ها ازدقت خوبی برخوردارند.

نتیجه‌گیری: در مقاله دیده می‌شود که تابع مفصل انتخاب‌شده نتایج قابل قبولی برای برآورد ذخیرهٔ خسارت‌های معوق داشته است.

توجه: مدت‌زمان بحث و انتقاد برای این مقاله تا ۱ اکتبر ۲۰۲۴ در وب‌سایت IJIR در «نمایش مقاله» باز است.

مقدمه

موضوع ذخائر و محاسبهٔ آن یکی از مسائل بسیار مهم شرکت‌های بیمه است و صنعت بیمهٔ کشور به‌دنبال روشی مناسب با شرایط بومی کشور برای حل این مسئله است. تاکنون این برآورد در سطح کلان و با روش نردبان زنجیری انجام شده است و ایرادهای زیادی ازسوی بیمهٔ مرکزی به آن وارد بوده است. هدف این مقاله ترویج و بومی‌سازی روش برآورد ذخیرهٔ معوق در سطح خرد با استفاده از تابع مفصل است.

برآورد ذخیرهٔ خسارت معوق متکی به پیش‌بینی میزان نهایی خسارت‌هایی است که هنوز تسویه نشده‌اند و بیمه‌گر متعهد به پرداخت آن است (Karimi, 2019). ممکن است خسارت‌ها در دورهٔ مالی مربوطه پرداخت نشوند، زیرا معمولاً خسارت‌های بزرگ، مدت‌زمان بیشتری برای تسویه شدن لازم دارند بنابراین با داده‌های سانسورشده مواجهیم (به‌عنوان یک منبع برای داده‌های سانسورشده می‌توان به Klein and Moeschberger (2003) مراجعه کرد). در واقع، در تاریخی معین، یا خسارتی تسویه شده است که در این صورت مقدار نهایی آن مشخص است، یا هنوز تسویه نشده است و در این صورت مبلغ نهایی خسارت سانسور شده است.

دو رویکرد کلی برای برآورد ذخیرهٔ خسارت معوق وجود دارد: رویکرد سطح کلان و رویکرد سطح خرد. در رویکرد سطح کلان معمولاً از روش نردبان زنجیری استفاده می‌شود که در Mack (1993)، Mack and Venter (2000)، Verrall (2000)، England and Verrall (2002) و Saluz *et al.* (2014) شرح داده شده است (که موضوع بحث این پژوهش نیست). در این مقاله از رویکرد سطح خرد استفاده می‌شود و برای پیش‌بینی مقدار خسارت معوق از مشخصات شناخته‌شدهٔ خسارت به‌عنوان متغیرهای پیش‌بینی‌کننده استفاده می‌شود. برآورد در سطح خرد، در عمل کارایی خوبی دارد، اما از نظر اجرایی پیچیده است (Hesselager, 1994). در رویکرد سطح خرد مبلغ هر خسارت تسویه‌نشده به‌صورت جداگانه برآورد می‌شود (Norberg, 1993; Norberg, 1999; Jin, 2017; Pigeon *et al.*, 2014) و مدت‌زمان پرداخت هر خسارت متغیر مهمی به شمار می‌رود (Yanez and Pigeon, 2021; Taha *et al.*, 2021; Lopez and Milhaud, 2021). در پژوهش‌هایی مانند رویکرد Jin (2017) و Ayuso and Santolino (2008)، متغیرهای مدت‌زمان پرداخت خسارت و مبلغ خسارت به‌طور جداگانه بررسی شده‌اند و Antonio *et al.* (2016) مدلی چندحالتی برای مبلغ خسارت در نظر گرفته‌اند.

رویکرد تابع مفصل برای تبیین ساختار وابستگی بین متغیرهای مدت‌زمان پرداخت خسارت و زمان وقوع خسارت را Zhao and Zhou (2010) مطرح کرد. ما در این مقاله از رویکرد Lopez *et al.* (2016) برای مدل‌سازی ساختار وابستگی متغیرهای مدت‌زمان پرداخت خسارت و مبلغ خسارت استفاده می‌کنیم. در این رویکرد اطلاعات در سطح خرد تحلیل می‌شوند. ابتدا مدل‌سازی (جداگانه) توزیع‌های حاشیه‌ای میزان خسارت و مدت‌زمان تسویه براساس متغیرهای کمکی انجام می‌شود، سپس این توزیع‌های حاشیه‌ای با یک تابع مفصل به یکدیگر پیوند داده می‌شوند و ذخیرهٔ خسارت‌های معوق با استفاده از تابع مفصل برآورد می‌شود. پس از این مقدمه، در بخش دوم این مقاله، فرضیات مسئله و روش‌های انتخاب توزیع‌های حاشیه‌ای و توزیع توأم تبیین می‌شود. در بخش سوم با استفاده از شبیه‌سازی کارایی مدل ساخته‌شده در بخش دوم ارزیابی می‌شود. در بخش چهارم مدل ساخته‌شده برای پیش‌بینی ذخیرهٔ خسارت‌های معوق یکی از شرکت‌های بیمه به کار برده خواهد شد.

مبانی نظری پژوهش

فرض کنید متغیرهای تصادفی M و T به‌ترتیب، نشان‌دهندهٔ مقدار خسارت و مدت‌زمان تسویهٔ خسارت، از زمان اعلام گزارش باشند. همچنین فرض کنید متغیر تصادفی D زمان بین اعلام گزارش تا تاریخ شروع مطالعه، و بردار $\mathbf{X}$، بردار متغیرهای پیش‌بینی‌کننده باشند. در این صورت تعریف می‌کنیم:

$$\begin{cases} Y = inf\left(T, D\right) \\ \delta = 1_{\mathrm{T \le D}} \\ N = \delta M \end{cases}$$

و برای تعداد n خسارت، مجموعه‌ای مستقل و هم توزیع $(\mathrm{Y_i, \delta_i, D_i, X_i, N_i})_{1 \le i \le n}$ از بردار تصادفی *(Y, δ, D, X, N)* به‌ازای $i = 1, \ldots, n$ هستند. در مدل سطح خرد، M_i مقدار خسارت شمارهٔ i است، بردار p بعدی، $X_i \in R^p$، بردار متغیرهای کمکی که خصوصیات هر خسارت است. T_i مدت‌زمان تسویهٔ خسارت است (یعنی زمان بین اعلام وقوع خسارت و تسویهٔ نهایی آن). این متغیر ممکن است با D_i سانسور شود. متغیر δ_i یک متغیر نشانگر است که برای موارد تسویه‌شده δ_i=1 و برای موارد تسویه‌نشده δ_i=0 است. همچنین متغیرهای $N_i = \delta_i M_i$ و $Y_i = inf(T_i, D_i)$ را تشکیل می‌دهیم. این مدل مشابه مدلی است که در Lopez *et al.* (2016) استفاده شده است، با این تفاوت که در اینجا فرض بر این است که $\mathrm{D_i}$ مشاهده شده است.

فرض می‌کنیم $(\mathrm{T, M, X})$ مستقل از D باشند. به‌عبارت‌دیگر طول بازهٔ زمانی اعلام خسارت تا شروع مطالعه، مستقل از متغیرهای کمکی X، مبلغ خسارت، طول دورهٔ زمانی پرداخت خسارت، است.

در بخش بعدی روش‌های برآورد توزیع‌های حاشیه‌ای $\mathrm{T \mid X = x}$ و $\mathrm{M \mid X = x}$ شرح داده می‌شوند.

برآورد توزیع‌های حاشیه‌ای

برای مدل‌سازی توزیع شرطی $\mathrm{T \mid X = x}$، از تابع نرخ خطر شرطی

$$h(t / x) = \lim_{dt \to 0^+} \frac{d\mathbb{P}(\mathrm{T \in [t,\ t + dt] \mid X = x})}{dt}$$

و مدل رگرسیون کاکس Cox and Oakes (1984)، به‌صورت $h(t|x) = h_0(t)\exp(a^T x)$ که در آن h_0 یک تابع پایهٔ ناشناخته و α بردار پارامتر نامعلوم است. برآورد α را می‌توان با استفاده از روش حداکثر شبه‌درستنمایی انجام داد، (در نظریهٔ آماری تابع شبه‌درستنمایی تقریبی برای تابع توزیع احتمال مشترک مجموعه‌ای از متغیرهای تصادفی است. کاربرد عملی آن این است که می‌تواند تقریبی برای تابع احتمال مجموعه‌ای از داده‌های مشاهده‌شده ارائه کند که ممکن است روش محاسباتی ساده‌تری برای تخمین ارائه کند، یا ممکن است راهی برای به دست آوردن برآوردهای صریح از پارامترهای مدل ارائه کند).

تابع خطر تجمعی $H_0(t) = \int_0^t h_0(s)\,ds$ را می‌توان با استفاده از برآوردگر برسلو Burr (1994) به‌صورت ناپارامتری برآورد کرد.

مدل‌سازی توزیع شرطی $M | X = x$ براساس مدل خطی تعمیم یافته است (Nelder and Baker, 1972; Madsen and Thyregod, 2010).

در این مدل یک تبدیل $\psi(M)$ شناخته شده از M در نظر گرفته می‌شود و فرض می‌شود $g(E[\psi(M)|X]) = \beta_0^T X$ که در آن g یک تابع پیوند شناخته شده است. برآورد پارامتر β_0 را می‌توان با استفاده از

$$\widehat{\beta} = \arg\max_{\beta} \sum_{i=1}^{n} W_{in} \log f_{\beta}\left(\psi(N_i)\middle|\beta^T X_i\right) \tag{۱}$$

که در آن $f_\beta(.|\beta^T x)$ چگالی شرطی از $\psi(M)|X$ با پارامتر β است، وزن‌ها

$$W_{in} = \frac{\delta_i}{n\hat{S}_C(Y_i)} \tag{۲}$$

هستند که در آنها $\hat{S}_C(t) = n^{-1}\sum_{i=1}^{n} 1_{c_{i\geq t}}$ تابع بقا تجربی و یک برآوردگر از $S_C(T) = \mathbb{P}(C \geq t)$ است. رابطهٔ (۱) و (۲) مربوط به برآورد حداکثر درستنمایی وزنی است که وزن‌ها برای اصلاح کاهش اریبی ناشی از سانسور اعمال شده‌اند.

مدل‌سازی براساس تابع مفصل

در مدل سطح خرد، متغیر T و متغیر M وابسته هستند. در واقع، انتظار می‌رود که خسارت‌هایی با مدت‌زمان طولانی T قبل از تسویه شدن، با مقدار بزرگی از M همراه باشند، زیرا به احتمال زیاد مطالبات کوچک به‌سرعت تسویه می‌شوند. مزیت مدل‌سازی ساختار وابستگی متغیر T و متغیر M از دیدگاه تابع مفصل در این است که می‌توان ساختار وابستگی را به‌صورت جداگانه از توزیع‌های حاشیه‌ای در نظر گرفت. بنابراین، دشواری یافتن خانواده‌ای از توزیع‌های پارامتری که با هر دو توزیع حاشیه‌ای سازگار باشد، مطرح نیست (Nelsen, 2006).

ایدهٔ اصلی تابع مفصل از قضیهٔ اسکلار است که بیان می‌کند برای دو متغیر تصادفی X و Y با تابع توزیع توأم $F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$ و توزیع‌های حاشیه‌ای F_Y، F_X وجود دارد تابع C به‌گونه‌ای که

$$F_{X,Y}(x,y) = C\left(F_X(x), F_Y(y)\right) \tag{۳}$$

تابع C یک تابع مفصل دومتغیره است (Sklar, 1959).

عکس قضیهٔ اسکلار بیان می‌کند که برای هر $(u, v) \in [0,1] \times [0,1]$ داریم

$$C(u, v) = F_{X,Y}\left(F_X^{-1}(u), F_Y^{-1}(v)\right) \tag{۴}$$

که در آن $F_X^{-1}(\cdot)$ تابع وارون F_X و $F_Y^{-1}(\cdot)$ تابع وارون F_Y است.

در حالت پارامتری تابع C متعلق به یک خانوادهٔ پارامتری از خانوادهٔ مفصل‌های $C = \{C_0 : \theta \in \Theta\}$ است که در آن $\Theta \subset \mathbb{R}^k$ و θ پارامتر وابستگی است.

تابع c_θ تابع چگالی C_θ است، که در صورت وجود از رابطهٔ $c_\theta(u,v) = \frac{\partial^2_{u,v} C_\theta(u,v)}{\partial_u \partial_v}$ به‌دست می‌آید.

برآوردگر حداکثر درستنمایی θ عبارت است از،

$$\hat{\theta}^{ML} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log c_\theta(u_i, v_i)$$

در جدول ۱ فرمول تابع توزیع و حدود پارامتر چند خانوادهٔ معروف از توابع مفصل بیان شده است.

ضریب همبستگی τ کندال با استفاده از تعریف تابع مفصل CC به‌صورت زیر است (Nelsen, 2006):

$$\tau_{X,Y} = \tau_C = 4\iint C(u,v)\,dC(u,v) - 1 \tag{۵}$$

همچنین ضرایب وابستگی دمی بالایی و پایینی، به‌ترتیب، عبارت‌اند از (Nelsen, 2006):

$$\lambda_u = \lim_{t\to 1^-} P\left(Y > F_Y^{-1}(t)\middle|X > F_X^{-1}(t)\right) = 2 - \lim_{t\to 1^-}\frac{1-C(t,t)}{1-t} \tag{۶}$$

$$\lambda_l = \lim_{t\to 0^+} P\left(Y \leq F_Y^{-1}(t)\middle|X \leq F_X^{-1}(t)\right) = \lim_{t\to 0^+}\frac{C(t,t)}{t} \tag{۷}$$

ضریب وابستگی دمی بالایی (پایینی) بیان می‌کند که چگونه مقادیر بزرگ (کوچک) یک متغیر با مقادیر کوچک (بزرگ) متغیر دیگر رخ می‌دهند. خانواده‌های توابع مفصل به ساختارهای وابستگی متفاوتی منجر می‌شوند، از جمله اینکه در خانواده‌های پلاکت، نرمال و فرانک به‌ازای تمام مقادیر مجاز پارامتر، ضرایب وابستگی دمی بالایی و پایینی صفر است. در خانواده‌های گامبل، جو و گالامبوس ضریب وابستگی دمی بالایی غیرصفر و ضریب وابستگی دمی پایینی صفر است.

جدول ۱: تابع توزیع و حدود پارامتر خانواده‌های پلاکت، فرانک،گامبل، نرمال، جو و گالامبوس
Table 1: Distribution functions and the range of their parameter of Plackett, Frank, Gumbel, Normal, Joe and Galambus families

خانواده	حدود پارامتر	تابع مفصل
پلاکت	(0,∞)-{1}	$C_\theta(u,v)=\frac{1+(\theta-1)(u+v)-\sqrt{(1+(\theta-1)(u+v))^2-4uv\theta(\theta-1)}}{2(\theta-1)}$
فرانک	[1,∞)	$C_\theta(u,v)=-\frac{1}{\theta}\ln(1+\frac{(e^{-u\theta}-1)(e^{-v\theta}-1)}{e^{-\theta}-1})$
گامبل	[-1,1]	$C_\theta(u,v)=\exp(-((-\ln u)^\theta+(-\ln v)^\theta)^{\frac{1}{\theta}}$
نرمال	[-1,1]	$C_\theta(u,v)=\Phi_\theta^{-1}(\Phi^{-1}(u),\Phi^{-1}(v))$
جو	[1,∞)	$C_\theta(u,v)=1-((1-u)^\theta+(1-v)^\theta-(1-u)^\theta(1-V)^\theta)^{\frac{1}{\theta}}$
گالامبوس	(0,∞)	$C_\theta(u,v)=uv\cdot\exp[(-\ln vu)^{-\theta}+(-\ln v)^{-\theta}]^{\frac{-1}{\theta}}$

برای متغیرهای تصادفی (T,M) اگر $F_{T,M|X}(t,m\mid x)$ تابع توزیع توأم آنها به شرط $X=x$ باشد، با توجه به (۳)، تابع C_x نشان‌دهندهٔ تابع مفصل توأم (T,M) به شرط $X=x$ است. در این صورت:

$$F_{T,M|X}(t,m|x)=C_{x,\theta}\left(F_{T|X}(t|x),F_{M|X}(m|x)\right) \quad (۸)$$

در اینجا فرض می‌کنیم فقط توزیع‌های حاشیه‌ای (T,M) به X بستگی دارند و ساختار وابستگی (T,M) به مقدار X بستگی ندارد، بنابراین $C_{x,\theta}=C_\theta$ است.

با فرض $U_i=F_{T|X}(T_i\mid X_i)$ و $V_i=F_{M|X}(M_i\mid X_i)$، متغیرهای U_i و V_i به‌طور یکنواخت در فاصلهٔ [0, 1] توزیع شده‌اند، و تابع توزیع توأم آنها C_θ است. معمولاً U_i و V_i را به دلیل اینکه توزیع‌های حاشیه‌ای شرطی $F_{T|X}$ و $F_{M|X}$ کاملاً مشخص نیستند و داده‌های سانسورشده وجود دارند، نمی‌توان به‌طور دقیق محاسبه کرد. در این صورت برآورد تجربی مقادیر $\hat{U}_i=\hat{F}_{T|X}(Y_i\mid X_i)$ و $\hat{V}_i=\hat{F}_{N|X}(N_i\mid X_i)$ محاسبه می‌شوند و برآوردگر پارامتر مفصل با توجه به برآوردگر حداکثر درستنمائی وزنی رابطهٔ (۱) به‌صورت زیر خواهد بود:

$$\hat{\theta}=\arg\max_{\theta\in\Theta}\sum_{i=1}^{n}\log c_\theta(\hat{u}_i,\hat{v}_i) \quad (۹)$$

که در آن W_{in} در (۲) تعریف شده است. در مورد داده‌های سانسورشده، یعنی وقتی $\delta_i=0$ آنگاه $W_{in}=0$ است، بنابراین W_{in} یک وزن برای اصلاح اریبی ناشی از سانسور است.

پیش‌بینی مقدار نهایی خسارت‌ها

در قسمت قبل توزیع توأم (T,M) به شرط $X=x$ با استفاده از تابع مفصل ساخته شد. از این مدل برای پیش‌بینی مقدار نهایی هر خسارت استفاده می‌شود. برای خسارت‌هایی که سانسور نشده‌اند تابع چگالی توأم شرطی با استفاده از توزیع توأم شرطی به‌دست می‌آید. اما برای یک خسارت تسویه‌نشده که برای آن $\delta_i=0$ است، معلوم است که متغیر مشاهده‌نشدهٔ T_i بزرگ‌تر از Y_i است، بنابراین، توزیع (T,M) به شرط $X=x_i$ و $T\geq Y_i$ مورد نیاز است که دارای برآورد تابع چگالی $f_{Y_i,X_i}(t,m)$ است.

از رابطهٔ (۸) تابع چگالی توأم (T,M) به شرط $X=x$ عبارت است از:

$$f_{T,M|X}(t,m|x)=c_{x,\theta}\left(F_{T|X}(t|x),F_{M|X}(m|x)\right)f_{T|X}(t|x)f_{M|X}(m|x)$$

بنابراین تابع چگالی توأم (T,M) به شرط $X=x$ و $T\geq Y_i$ عبارت است از:

$$f_{T,M|X=x,T\geq Y_i}(t,m)=\frac{c_{x,\theta}\left(F_{T|X}(t\mid x),F_{M|X}(m\mid x)\right)f_{T|X}(t|x)f_{M|X}(m\mid x)(1_{t\geq Y_i})}{\int_{u=Y_i}^{\infty}\int_{v=-\infty}^{\infty}c_{x,\theta}\left(F_{T|X}(u\mid x),F_{M|X}(v\mid x)\right)f_{T|X}(u|x)f_{M|X}(v\mid x)\,du\,dv}$$

که به‌صورت زیر برآورد می‌شود:

$$\hat{f}_{Y_i,X_i}(t,m)=\frac{c_{\hat{\theta}}\left(\hat{F}_{T|X}(t\mid x),\hat{F}_{M|X}(m\mid x)\right)\hat{f}_{T|X}(t|x)\hat{f}_{M|X}(m\mid x)1_{t\geq Y_i}}{\int_{u=Y_i}^{\infty}\int_{v=-\infty}^{\infty}c_{\hat{\theta}}\left(\hat{F}_{T|X}(u\mid x),\hat{F}_{M|X}(v\mid x)\right)\hat{f}_{T|X}(u|x)\hat{f}_{M|X}(v\mid x)\,du\,dv} \quad (۱۰)$$

که در آن $\hat{f}_{T|X}(.\mid x)(\hat{f}_{M|X}(.|x))$ تابع چگالی شرطی برآوردشده $T\mid X=x$ و $(M\mid X=x)$ و $\hat{F}_{T|X}(.\mid x)(\hat{F}_{M|X}(.\mid x))$ تابع توزیع شرطی برآوردشده را نشان می‌دهد. سپس پیش‌بینی مبلغ خسارت (به شرط آنکه تا زمان Y_i پرداخت نشده باشد) براساس برآورد امید ریاضی خسارت از رابطهٔ

$$\hat{E}(M|X=x,T\geq Y_i)=\int_{m=0}^{+\infty}\int_{t=Y_i}^{\infty}m\hat{f}_{Y_i,X_i}(t,m)\,dt\,dm \quad (۱۱)$$

انجام می‌شود.

شبیه‌سازی

برای بررسی اعتبار روش پیش‌بینی یادشده در بخش ۲، با مقادیر پارامترهای خلاصه‌شده در جدول ۲، و سه خانوادهٔ مفصل پلاکت،

جدول ۲: پارامترهای شبیه‌سازی
Table 2: Simulation parameters

نام پارامتر	مقدار پارامتر	درصد مشاهدات سانسورشده
$\alpha_{0.1}$	0	
$\alpha_{0.2}$	0.5	
$\alpha_{0.3}$	0.2	
$\alpha_{0.4}$	-0.1	
$\beta_{0.0}$	0.1	
$\beta_{0.1}$	0.03	
$\beta_{0.2}$	0.05	
$\beta_{0.3}$	-0.02	
(k, λ)	(2.7,3)	30
(k, λ)	(2.5,3)	50
(k, λ)	(0.15,3)	60

جدول۳: اندازهٔ خطا به‌ازای $\tau = 0.25$ و $n = 5000$
Table 3: Error measurement for $\tau = 0.25$ and n=5000

درصد سانسور	خانوادهٔ مفصل	E_1	E_{1r}	E_2
30	پلاکت	17.61	0.21	0.014
	فرانک	20.61	0.27	0.012
	نرمال	28.88	0.30	0.018
50	پلاکت	15.27	0.15	0.014
	فرانک	12.93	0.27	0.013
	نرمال	19.41	0.19	0.012
60	پلاکت	14.04	0.18	0.033
	فرانک	11.04	0.29	0.039
	نرمال	17.27	0.26	0.024

فرانک و نرمال، شبیه‌سازی به شرح زیر انجام می‌شود:

مرحلهٔ ۱. به‌ازای $i = 1, \dots, n$ بردارهای $X_i = (X_{i1}, X_{i2}, X_{i3})'$ شبیه‌سازی می‌شود که در آن X_{ij}ها به‌طور مستقل از توزیع یکنواخت در فاصلهٔ (۱ و ۰) تولید می‌شوند.

مرحلهٔ ۲. به‌ازای $i = 1, \dots, n$ متغیر T_i از توزیع لگ - نرمال به صورت شرطی بر روی متغیرهای X_i شبیه‌سازی می‌شود.

مرحلهٔ ۳. به‌ازای $i = 1, \dots, n$ سانسور D_i به‌طور مستقل از توزیع وایبل با پارامترهای k و λ یعنی با تابع چگالی $f_k(d) = k\lambda^{-k}d^{k-1}\exp(-d/\lambda)^k 1_{d\geq 0}$ شبیه‌سازی می‌شود.

مرحلهٔ ۴. برای هر نمونهٔ شبیه‌سازی‌شده $(Y_i, N_i, D_i, \delta_i, X_i)_{1\le i\le n}$ برآورد تابع چگالی از رابطهٔ (۱۰) محاسبه می‌شود.

مرحلهٔ ۵. به‌ازای $i = 1, \dots, n$ پیش‌بینی $\hat{M}_i$ از رابطهٔ (۱۱) محاسبه می‌شود.

مرحلهٔ ۶. مراحل ۱ تا ۵ را به تعداد $B = 10000$ بار تکرار می‌کنیم تا به‌ازای $b = 1, \dots, B$ نمونه‌های $(Y_i^b, N_i^b, D_i^b, \delta_i^b, X_i^b)_{1\le i\le n}$ شبیه‌سازی شوند و مقادیر برآوردشده $\hat{M}_i^b$ به دست آیند.

برای تحلیل تأثیر قدرت سانسور، درصدهای مختلفی برای سانسور شدن در نظر گرفته می‌شود. اگر $R^b = \sum_{i=1}^n (1-\delta_i) M_i^b$ ذخیرهٔ مربوط به تمام مطالبات خسارت‌ها در مرحلهٔ bام باشد و $\hat{R}^b = \sum_{i=1}^n (1-\delta_i)\hat{M}_i^b$ مقدار پیش‌بینی‌شدهٔ آن باشد، خطاهای مختلف اندازه‌گیری که به شرح زیر تعریف شده‌اند، محاسبه و نتایج در جدول ۳ ثبت شد.

$$E_1 = \frac{1}{B}\sum_{b=1}^{B}\left|R^b - \hat{R}^b\right|$$

$$E_{1r} = \frac{1}{B}\sum_{b=1}^{B}\left|\frac{R^b - \hat{R}^b}{R^b}\right|$$

$$E_2 = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{\sum_{i=1}^n (1-\delta_i)}\sum_{i=1}^{n}(1-\delta_i)\left|\hat{M}_i^b - M_i^b\right|$$

خطای E_1 نشان‌دهندهٔ میانگین قدر مطلق تفاضل مقدار واقعی ذخیره و برآورد آن است. خطای E_{1r} میانگین خطای نسبی برآورد را اندازه می‌گیرد. خطای E_2 میانگین قدر مطلق تفاضل برآورد هر خسارت پرداخت‌نشده و مقدار واقعی خسارت است.

جدول ۳ نشان می‌دهد برای هر مدل، خطای نسبی E_{1r} تحت توابع مفصل مختلف نسبتاً پایاست. وقتی نسبت سانسور افزایش می‌یابد E_1 تمایل به کاهش دارد زیرا تعداد خسارت‌های

جدول ۴: آمار توصیفی T (بر حسب روز) و M' (بر حسب ده میلیون ریال) قبل از حذف اثر تورم
Table 4: Descriptive statistics for T (in terms of day) and *M′ (in terms of ten million Rials) before removing the inflation*

متغیر	مینیمم	چارک اول	میانه	میانگین	چارک سوم	ماکزیمم
M'	0.0027	1.68	5.92	22.14	21.22	715.50
T	1	85	319	543	828	3234

پیش‌بینی‌شده افزایش می‌یابد و ازآنجاکه برآوردگرها به‌طور مجانبی ناریب هستند، بیش‌برآوردی‌ها را کم‌برآوردی‌ها خنثی می‌کنند. خطای E_2 (در مورد برآورد پارامتر مفصل) با نسبت سانسور افزایش می‌یابد، زیرا اطلاعات کمتر، به مدل با عملکرد ضعیف‌تری منجر می‌شود و تأثیر منفی بر پیش‌بینی دارد.

مطالعهٔ موردی دربارهٔ داده‌های واقعی

پژوهشکدهٔ بیمه یک مجموعه داده با ۲۷۰۲ مورد اعلام‌شده ناشی از مسئولیت حرفه‌ای کارفرما در یک دورهٔ ۸ ساله در اختیار ما قرار داد که شامل ۲۶۶۰ مورد خسارت بدون سانسور هستند که مبلغ نهایی آنها مشخص است و ۴۲ مورد سانسور شده است (آنهایی که در پایان این دورهٔ ۸ ساله تسویه نشده بودند). شایان ذکر است که به‌دلیل نبود داده‌های مناسب که بتوان آنها را به‌عنوان متغیرهای کمکی در نظر گرفت، در این مطالعهٔ موردی از متغیرهای کمکی صرف‌نظر شد. هدف برآورد مبلغ خسارت این ۴۲ مورد و برآورد ذخیره معوق کل است.

برای اجتناب از عوامل خارج از کنترل برای تجزیه‌وتحلیل داده‌ها با روشی که در ادامه شرح داده می‌شود، عامل تورم حذف می‌شود. متغیر T (برحسب روز) مدت‌زمان بین اعلام خسارت و تسویه‌حساب است. متغیر M' (برحسب ده میلیون ریال) مبلغ پرداخت‌شده ازسوی شرکت بیمه است. جدول ۴ آمارهای توصیفی مربوط به M', T را برای موارد تسویه‌شده، نشان می‌دهد.

حذف اثر تورم

از روش‌های مختلفی برای اصلاح اثر تورم بر مبلغ خسارت استفاده می‌شود. در یکی از ساده‌ترین روش‌ها با توجه به اینکه تورم باعث افزایش قیمت‌ها می‌شود، فرض می‌شود افزایش سالیانهٔ قیمت ثابت باشد و لگاریتم نسبت قیمت‌ها در دو سال متوالی ضریبی از سال فرض شده است، به‌عبارت‌دیگر:

$$\log \frac{M_i'}{M_i} = \mu . g_i$$

که در آن M_i' مقدار خسارت را قبل از حذف تورم، M_i مقدار خسارت بعد از حذف تورم، g_i سالی است که خسارت iام در آن اتفاق افتاده است و فرض می‌کنیم $g_i \in \{0,1,2,\ldots,G\}$ و ضریب μ ضریبی ثابت است که با استفاده از مجموعه داده‌ها برآورد می‌شود. در این مدل فرض می شود $(M_i)_{1\le i\le n}$ مستقل از $(g_i)_{1\le i\le n}$ است. و بنابراین از مدل زیر استفاده می‌شود.

$$\log M_i' = \mu . g_i + \log M_i \qquad (۱۲)$$

برای برآورد μ روش زیر به کار می‌رود:

- قرار می‌دهیم m_{ij} را به‌عنوان میانگین تعداد n_{ij} خسارت‌هایی که در سال مالی g_i رخ داده است و پس از j سال تسویه شده است،
- از رابطهٔ $\log m_{ij} \approx \mu g_i + \lambda_j$ که در آن $\lambda_j = E[\log M_i \mid T_i = j]$ برای هر j است، برآوردگر وزنی حداقل توان دوم خطا را برای μ_j محاسبه می‌کنیم، یعنی:

$$\arg\min_{a,b} \sum_{i: g_i + j \le G} n_{ij} \left(\log m_{ij} - a - b g_i\right)^2$$

- برآوردگر نهایی μ عبارت است از:

$$\hat{\mu} = \frac{\sum_j n_j^{1/2} \hat{\mu}_j}{\sum_j n_j^{1/2}} \qquad (۱۳)$$

که در آن $n_j = \sum_i n_{ij}$.

در نهایت، برای هر i که $\delta_i = 1$ است مقدار خسارت پس از حذف اثر تورم به صورت $\log \hat{M}_i = \log M_i' - \hat{\mu} g_i$ محاسبه می‌شود. جدول ۵ آمار توصیفی متغیر M بعد از حذف اثر تورم را نشان می‌دهد.

نمودارهای ۱ و ۲ هیستوگرام مربوط به مبالغ خسارات پرداخت‌شده (بعد از حذف اثر تورم) و هیستوگرام مدت‌زمان پرداخت خسارات را نشان می‌دهند.

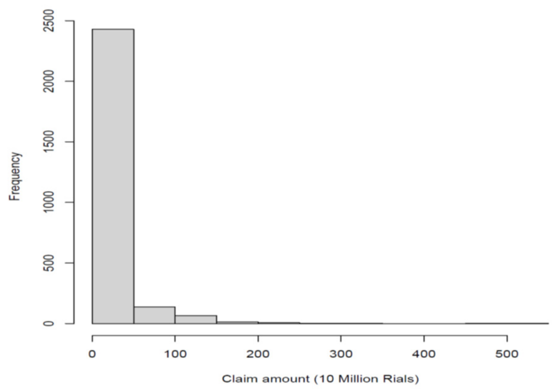
برآورد توزیع‌های حاشیه‌ای

با توجه به نمودارهای هیستوگرام متغیرهای T و M که حاکی از چولگی مثبت تابع چگالی هر دو متغیر است، توزیع‌های نمایی، گاما، لگ نرمال و وایبل را برای برازش به آنها آزمون کردیم، ولی هیچ‌یک از آنها برازش مناسبی نداشتند، لذا تصمیم گرفتیم از تقریب توزیع نرمال استفاده کنیم. مناسب‌ترین تبدیل برای متغیر M تبدیل معروف *Box-Cox* با ضریب $\lambda = 0.0554$ به‌صورت $M^* = h_1(m) = \frac{m^{\lambda}-1}{\lambda}, \lambda \neq 0$ و مناسب‌ترین تبدیل برای متغیر T تبدیل معروف *Order – Norm* به‌صورت $T^* = h_2(t) = \frac{\phi^{-1}\left(rank(t) - 0.5\right)}{n}$ تشخیص داده شد. جدول ۶ نتایج مربوط به آزمون نرمالیتی را برای متغیرهای T^* و M^* نشان می‌دهد.

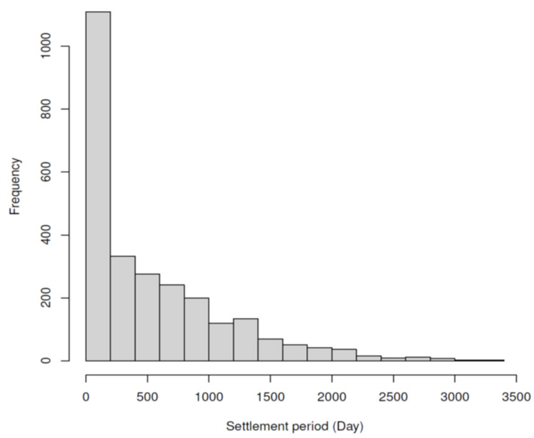
نمودارهای ۳ و ۴ هیستوگرام مربوط T^* و M^* هستند.

جدول ۵: آمار توصیفی M (برحسب ده میلیون ریال) بعد از حذف اثر تورم
Table 5: Descriptive statistics and *M (in terms of ten million Rials) after removing the inflation*

متغیر	مینیمم	چارک اول	میانه	میانگین	چارک سوم	ماکزیمم
M	0.0022	1.15	4.07	16.03	15.20	517.33



نمودار ۱: هیستوگرام مبلغ خسارت.
Fig. 1: Histogram of the claim amount



نمودار ۲: هیستوگرام مدت‌زمان پرداخت مبلغ خسارت.
Fig. 2: Histogram of the duration of the payment of claim

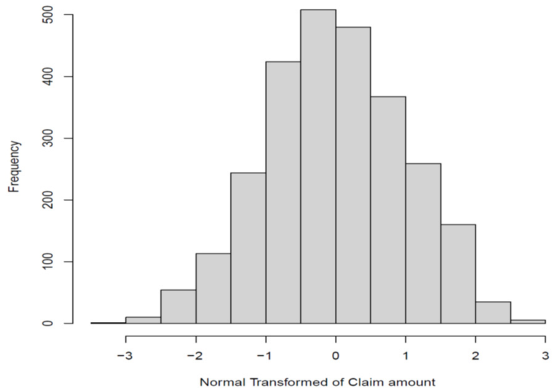
جدول ۶: نتایج مربوط به آزمون نرمالیتی را برای متغیرهای T^* و M^*
Table 6: Results of the Normality test for T^* and M^*

متغیر	*آمارهٔ آزمون کلموگروف ـ اسمیرنوف*	مقدار احتمال
M^*	0.0168	0.4382
T^*	0.0101	0.9469

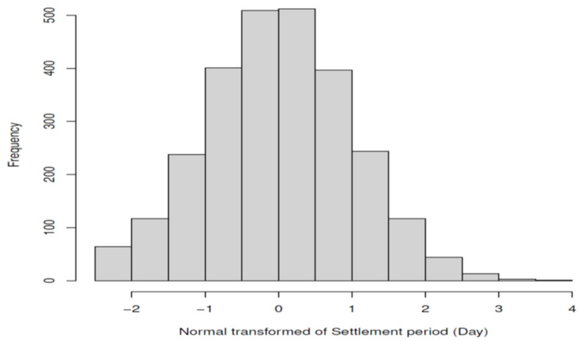
برای بیان وابستگی مبلغ خسارت پرداخت‌شده (پس از تبدیل به نرمال) در مقابل مدت‌زمان پرداخت خسارت (پس از تبدیل به نرمال) نمودار پراکنش آنها رسم شده است نمودار ۵ و همچنین برآورد نمونه‌ای ضریب وابستگی τ کندال و ضرایب وابستگی دمی بالایی و پایینی را برای داده‌های متغیرهای T^* و M^* محاسبه شده‌اند که در جدول ۷ نشان داده شده است.

برازش تابع مفصل

تابع مفصل تحت تبدیل‌های اکیداً افزایشی، پایاست بنابراین تابع مفصل متغیرهای T^* و M^* همان تابع مفصل متغیرهای (T,M) است. از بین توابع مفصل مختلفی که به داده‌های زوج (T^*,M^*) برازش داده شد، براساس آزمون نیکویی برازش، فرضیهٔ صفر برازش چهار تابع مفصل پلاکت، فرانک، نرمال و جو در سطح



نمودار ۳: هیستوگرام مبلغ خسارت پس از تبدیل به نرمال.
Fig. 3: Histogram of the claim amount after Normal transformation



نمودار ۴: هیستوگرام مدت‌زمان پرداخت مبلغ خسارت پس از تبدیل به نرمال.
Fig. 4: Histogram of the duration of the payment of claim after Normal transformation

معناداری 0.01 رد نشد. جدول ۸، مقدار احتمال آزمون نیکویی برازش، مقدار لگاریتم تابع درست‌نمایی، برآورد پارامتر و برآورد ضریب τ کندال برای هریک از این توابع مفصل را نشان می‌دهد.

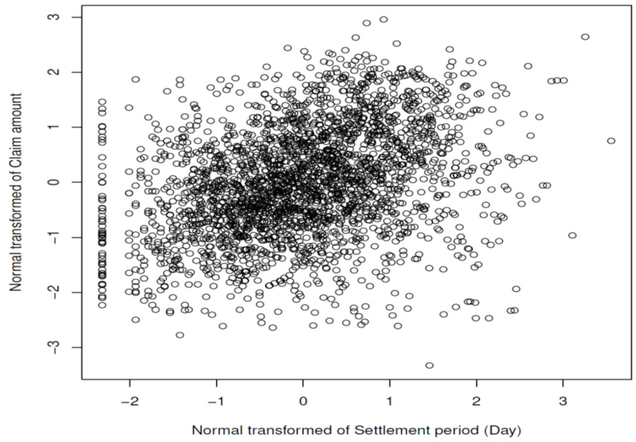
برآورد نمونه‌ای ضرایب وابستگی دمی براساس داده‌ها با هم برابر و تقریباً برابر صفر است، بنابراین خانواده‌های پلاکت، نرمال و فرانک گزینه‌های مناسبی هستند و تابع مفصل جو از این نظر مناسب نیست. از بین این سه خانواده براساس ملاک مقدار تابع درست‌نمایی، چون لگاریتم تابع درست‌نمایی تابع مفصل پلاکت از بقیه بیشتر است تابع مفصل پلاکت را به‌عنوان گزینهٔ نهایی تابع مفصل برای زوج (T^*, M^*) انتخاب می‌کنیم.

برآورد ذخیرهٔ خسارت معوق

با برازش تابع مفصل پلاکت به (T^*, M^*) با استفاده از رابطهٔ (۱۰) تابع چگالی توأم به‌دست می‌آید و برآورد خسارت مواردی که تسویه نشده‌اند با استفاده از رابطهٔ (۱۱) انجام می‌شود. تبدیل وارون باکس کاکس روی مقادیر $\hat{M}^*$ اعمال شده و در نهایت برآوردی از مقدار ذخیره به‌صورت $\hat{R}=\sum_{i=1}^{n}(1-\delta_i)\hat{M}_i$ به دست می‌آید. در داده‌های مورد مطالعه ۴۲ مورد تسویه‌نشده در بین ۲۷۰۲ خسارت اعلام‌شده وجود داشت که برآورد شده و نتایج در جدول ۹ نشان داده شده است.

برای ارزیابی خطای برآورد با این روش در این مجموعه داده، از بین داده‌هایی که مبلغ خسارت آنها معلوم است و تسویه شده به تعداد ۱۰۰۰ بار نمونه‌هایی به حجم ۴۰ (هر بار فرض شده که ۴۰ مورد تسویه‌نشده وجود دارد) شبیه‌سازی کردیم. در این نمونه‌ها مبلغ خسارت معلوم است، ولی این مبالغ برآورد نیز شده‌اند و خطاهای E_1 ، $E_{1.r}$ ، E_2 محاسبه و نتایج در جدول ۹ نشان داده شده است.

نتایج جدول ۱۰ حاکی از آن است که در ۱۰۰۰ بار شبیه‌سازی نمونه‌های ۴۰تایی از خسارت‌های معلوم، میانگین مقدار واقعی کل خسارت پرداخت‌شده ۴۰ مورد برابر ۴۲۱/۲۱ (ده میلیون ریال) است و با فرض اینکه برای این نمونه‌ها مبلغ خسارت سانسور شده است، میانگین برآورد کل خسارت برابر ۴۹۸٫۹۱ (ده میلیون ریال)، متوسط قدر مطلق خطای برآورد خسارت کل برابر ۲۶/۳۹ (ده میلیون ریال)، متوسط خطای نسبی برآورد خسارت کل برابر ۰/۱۲۹ و میانگین میانگین قدر مطلق هر مورد خطا برابر ۰/۷۹ (ده میلیون ریال) است. در نمودار ۶ هیستوگرام فراوانی برآورد ذخیرهٔ کل به‌دست‌آمده در



نمودار ۵: نمودار پراکنش مبلغ خسارت در مقابل مدت‌زمان پرداخت خسارت پس از تبدیل به نرمال.
Fig. 5: Scatter plot of claim amount versus the duration of its payment after Normal transformation

جدول ۷: برآورد نمونه‌ای از ضریب وابستگی τ کندال و ضرایب وابستگی دمی بالایی و پایینی
Table 7: Sample estimations of the Kendall's τ and upper and lower tail dependence coefficients

زوج متغیر	τ	λ_l	λ_u
T^*, M^*	0.25	0.05	0.05

جدول ۸: نتایج آزمون نیکویی برازش و برآورد پارامترهای توابع مفصل منتخب
Table 8: Results of the good ness of fit test and the parameter estimations of the selected copula function

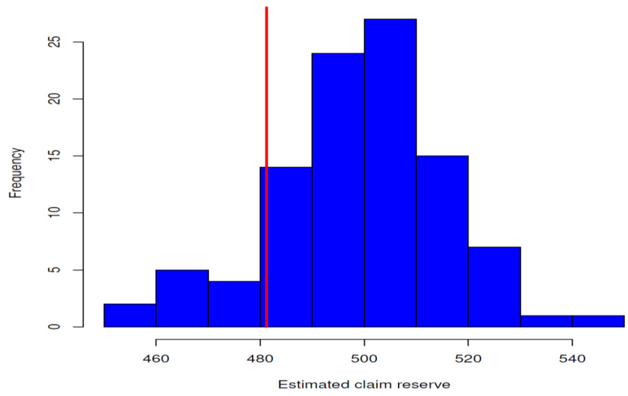
خانوادهٔ مفصل	مقدار احتمال	لگاریتم درست‌نمایی	برآورد پارامتر	برآورد τ	برآورد (λ_l, λ_U)
پلاکت	**0.0528**	**5.92**	**0.0762**	**0.2455**	**(0,0)**
نرمال	0.0544	5.45	0.3455	0.2232	(0,0)
فرانک	0.0244	5.23	2.1043	0.2241	(0,0)
جو	0.1043	7.82	1.4791	0.2122	(0,0.04)

جدول ۹: برآورد مبلغ خسارت‌های تسویه‌شده (برحسب ده میلیون ریال)
Table 9: Estimations of unsettled claim amount (in terms of *ten million* Rials)

ردیف	برآورد مبلغ	ردیف	برآورد مبلغ	ردیف	برآورد مبلغ	ردیف	برآورد مبلغ
10	8.63	772	14.79	1766	10.65	2680	9.04
166	10.57	940	9.28	2660	8.87	2683	7.99
167	10.64	941	13.76	2665	10.53	2684	9.13
345	17.59	963	13.40	2666	8.77	2685	8.57
445	12.12	1453	12.51	2668	11.28	2687	9.46
446	12.20	1454	12.13	2671	9.71	2692	10.03
569	14.16	1455	9.67	2672	10.28	2696	9.49
648	11.38	1456	12.85	2674	9.31	2698	9.93
649	11.47	1457	13.45	2675	9.51	2700	9.94
770	14.37	1514	9.67	2676	10.01		جمع
771	14.50	1595	11.03	2678	10.74		$\hat{R}$= 463.50

جدول ۱۰: متوسط مقدار واقعی خسارات، متوسط مقدار برآوردشدهٔ ذخیره و خطاهای برآورد در ۱۰۰۰ بار شبیه‌سازی
Table 10: The average of real amount of claims, the average of estimated amount of claims and estimation errors in 1000 times of simulation

$\bar{R}$	$\bar{\bar{R}}$	E_1	E_{1r}	E_2
481.21	498.91	26.39	0.129	0.79



نمودار ۶: فراوانی برآورد ذخیره، خط قرمز نشان‌دهندهٔ میانگین واقعی ذخیره
Fig. 6: Reserve estimate frequency, red line represents the actual average reserve

شبیه‌سازی و میانگین مقدار واقعی ذخیره (خط قرمزرنگ) نشان داده شده است.

جمع بندی و پیشنهادها

در این مقاله، برای پیش‌بینی مبلغ نهایی یک خسارت تسویه‌نشده، از ساختار وابستگی بین مقدار خسارت و طول عمر آن استفاده شد. این ساختار وابستگی با تابع مفصل که امکان توصیف وابستگی را به‌صورت جداگانه‌ای از مدل توزیع‌های حاشیه‌ای فراهم می‌کند، تبیین شد.

برای بهتر شدن دقت پیش‌بینی در صورت موجود بودن خصوصیات شناخته‌شده‌ای از مبلغ خسارت و طول عمر آن (به‌عنوان متغیرهای کمکی) می‌توان مدل‌سازی حاشیه‌ای را براساس توزیع شرطی به این خصوصیات انجام داد. برای انتخاب تابع مفصل مناسب به زوج متغیر مبلغ خسارت و طول عمر، از تطابق برآورد ضرایب وابستگی دمی نمونه‌ای (ناپارامتری) و برآورد ضرایب وابستگی دمی تابع مفصل استفاده کردیم. از بین توابع مفصل‌های انتخاب‌شده در مرحلهٔ تطابق، با استفاده از آزمون نیکویی برازش توابعی که در سطح معناداری ۰,۰۱ رد نمی‌شدند گزینش شدند و سپس با معیار بیشترین مقدار تابع درستنمایی، تابع مفصل پلاکت که بیشترین مقدار درستنمایی را داشت به‌عنوان تابع مفصل منتخب برگزیده شد. پس از انتخاب تابع مفصل با استفاده از قضیهٔ اسکلار تابع توزیع و تابع چگالی دومتغیره مبلغ خسارت و مدت‌زمان پرداخت به دست آمد. ازآنجاکه برای هر خسارت تسویه‌نشده، مدت‌زمان پرداخت از زمان شروع مطالعه بیشتر است تابع چگالی شرطی مبلغ آن خسارت به شرط آنکه مدت‌زمان پرداخت آن از یک مدت مشخص بیشتر باشد، محاسبه شد. سپس مبلغ هر خسارت تسویه‌نشده با استفاده از تقریب میانگین شرطی آن پیش‌بینی شد و جمع مقادیر این پیش‌بینی‌ها به‌عنوان پیش‌بینی کل خسارت معوق در نظر گرفته شد. همان‌گونه که شرح داده شد در این روش انتخاب تابع مفصل و انتخاب متغیرهای کمکی اهمیت بسیار زیادی در دقت پیش‌بینی دارند.

شایان ذکر است که استفاده از این روش ملزومات خاصی نیاز ندارد و هزینهٔ اضافی به سیستم وارد نمی‌کند. فقط کافی است اطلاعات مورد نیاز با دقت ثبت و نگهداری شوند.

مشارکت نویسندگان

صدیقه شمس: طرح موضوع، ارائهٔ راه‌حل، تدوین مقاله، ارائهٔ الگوریتم محاسباتی، تنظیم کدهای برنامهٔ کامپیوتری، مریم اثنی‌عشری: پردازش مسئله، مشاوره در بسط راه‌حل، یافتن داده‌ها، کمک در تدوین مقاله، محبوبه پیاده کوهسار: پالایش و تنظیم داده‌ها، کمک در نوشتن کدهای برنامه کامپیوتری.

تشکر و قدردانی

از پیشنهادهای داوران محترم که به غنای علمی مقاله کمک کردند، بسیار سپاسگزاریم.

تعارض منافع

نویسنده (گان) اعلام می‌دارند که در خصوص انتشار این مقاله تضاد منافع وجود ندارد. علاوه‌براین، موضوعات اخلاقی شامل سرقت ادبی، رضایت آگاهانه، سوءرفتار، جعل داده‌ها، انتشار و ارسال مجدد و مکرر ازسوی نویسندگان رعایت شده است.

دسترسی آزاد

یادداشت ناشر

ناشر نشریهٔ پژوهشنامهٔ بیمه با توجه به مرزهای حقوقی در نقشه‌های منتشرشده بی‌طرف باقی می‌ماند.

منابع

Antonio, K.; Godecharle, E.; Van Oirbeek, R., (2016). A multi-state approach and flexible payment distributions for micro-level reserving in general insurance. SSRN., 1-35 **(35 Pages)**.

Ayuso, M.; Santolino, M., (2008). Prediction of individual automobile RBNS claim reserves in the context of Solvency II. IREA–Working. Pap.

Burr, D., (1994). On inconsistency of Breslow's estimator as an estimator of the hazard rate in the Cox model. Biom., 50(4): 1142-1145 **(4 Pages)**.

Cox, D.R.; Oakes, D., (1984). Analysis of survival data. CRC press.

England, P.D.; Verrall, R.J., (2002). Stochastic claims reserving in general insurance. Br. Actuarial. J., 8(3): 443-518 **(76 Pages)**.

Hesselager, O., (1994). A Markov model for loss reserving. ASTIN. Bull. J. IAA., 24(2): 183-193 **(11 Pages)**.

Jin, X., (2017). Micro-level loss reserving models with applications in workers compensation insurance. Univ. Wisconsin. Empirical. Pap., 1-38 **(38 Pages)**.

Karimi, A., (2019). Insurance generalities. Insurance research center (IRC) press., [In Persian]

Klein, J.P.; Moeschberger, M.L., (2003). Survival analysis: Techniques for censored and truncated data. Springer.

Lopez, O.; Milhaud, X., (2021). Individual reserving and nonparametric estimation of claim amounts subject to large reporting delays. Scand. Actuarial. J., 2021(1): 34-53 **(20 Pages)**.

Lopez, O.; Milhaud, X.; Thérond, P.E., (2016). Tree-based censored regression with applications in insurance. Electr. J. Stat., 10(2): 2685-2716 **(32 Pages)**.

Mack, T., (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN. Bull. J. IAA., 23(2): 213-225 **(13 Pages)**.

Mack, T.; Venter, G., (2000). Comments on a comparison of stochastic models that reproduce chain ladder reserve estimates. Insur. Math. Econ., 26(1): 109-109 **(1 Pages)**.

Madsen, H.; Thyregod, P., (2010). Introduction to general and generalized linear models. CRC press.

Nelder, J.A.; Baker, R.J., (1972). Generalized linear models. Wiley online library.

Nelsen, R.B., (2006). An introduction to copulas. Springer.

Norberg, R., (1993). Prediction of outstanding liabilities in non-life insurance1. ASTIN. Bull. J. IAA., 23(1): 95-115 **(21 Pages)**.

Norberg, R., (1999). Prediction of outstanding liabilities II. Model variations and extensions. ASTIN. Bull. J. IAA., 29(1): 5-25 **(21 Pages)**.

Pigeon, M.; Antonio, K.; Denuit, M., (2014). Individual loss reserving using paid–incurred data. Insur. Math. Econ., 58: 121-131 **(11 Pages)**.

Saluz, A.; Bühlmann, H.; Gisler, A.; Moriconi, F., (2014). Bornhuetter-Ferguson reserving method with repricing. SSRN., 1-31 **(31 Pages)**.

Sklar, M., (1959). Fonctions de répartition à n dimensions et leurs marges. Ann. l'isup., 8(3): 229-231 **(3 Pages)**.

Taha, A.; Cosgrave, B.; Rashwan, W.; McKeever, S., (2021). Insurance reserve prediction: Opportunities and challenges. IEEE., 290-295 **(6 Pages)**.

Verrall, R.J., (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. Insur. Math. Econ., 26(1): 91-99 **(9 Pages)**.

Yanez, J.S.; Pigeon, M., (2021). Micro-level parametric duration-frequency-severity modeling for outstanding claim payments. Insur. Math. Econ., 98: 106-119 **(14 Pages)**.

Zhao, X.; Zhou, X., (2010). Applying copula models to individual claim loss reserving methods. Insur. Math. Econ., 46(2): 290-299 **(10 Pages)**.

AUTHOR(S) BIOSKETCHES — معرفی نویسندگان

صدیقه شمس، آمار، دانشیار گروه آمار، دانشکده علوم ریاضی، دانشگاه الزهرا(س)، تهران، ایران

- Email: *s.shams@alzahra.ac.ir*
- ORCID: 0000-0002-8831-3372
- Homepage: http://staff.alzahra.ac.ir/shams

مریم اثنی‌عشری، استادیار گروه پژوهشی بیمه‌های اموال و مسئولیت، پژوهشکده بیمه، تهران، ایران

- Email: *esnaashari@irc.ac.ir*
- ORCID: 0000-0001-5337-9665
- Homepage: https://www.irc.ac.ir/fa-IR/Irc/5020/page

فاطمه پیاده کوهسار، دانش آموخته کارشناسی ارشد رشته آمار ریاضی گروه آمار، دانشکده علوم ریاضی، دانشگاه الزهرا(س)، تهران، ایران

- Email: *mahbobehpiadehkohsar@gmail.com*
- ORCID: 0009-0006-4658-8622
- Homepage: https://math.alzahra.ac.ir/page-Statistics/fa

HOW TO CITE THIS ARTICLE

Shams, S.; Esna-Ashari, M.; Piadeh Kouhsar, M., (2024). Estimation of micro-level claim reserving using Copula function. Iran. J. Insur. Res., 13(3): 275-286.

DOI: 10.22056/ijir.2024.03.06

URL: https://ijir.irc.ac.ir/article_160314.html?lang=en